

Section 10.1

We now learn about contingency tables.

Definition: A **contingency table** (or **two-way frequency table**) is a table in which frequencies correspond to two variables. (One variable is used to categorize rows, and a second variable is used to categorize columns.)

Here is an example:

A 1992 poll conducted by the University of Montana classified respondents by gender and political party, as shown in the table below

	Democrat	Republican	Independent
Male	36	45	24
Female	48	33	16

Definition: A **test of independence** tests the null hypothesis that in a contingency table, the row and column variables are independent.

Here is the test statistic for a Test of Independence:

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

where O is the observed frequency in a cell and E is the expected frequency found by evaluating:

$$E = (\text{row total})(\text{column total})/(\text{grand total}).$$

The degrees of freedom = $(r-1)(c-1)$, where r is the number of rows and c is the number of columns.

Here are the requirements:

1. The sample data are randomly selected.
2. The sample data are represented as frequency counts in a two-way table.
3. For every cell in the contingency table, the *expected* frequency E is at least 5.

Exercise 1. We are interested in seeing in whether giving flu-shots reduces the likelihood of getting the flu.

Result	Flu shot	No flu shot
Got the flu	54	157
Did not get the flu	196	103

(a) Calculate the marginal frequencies.

(b) Find the expected frequency for each cell in the contingency table. Assume that the variables are independent.

Class Exercise 1. Does the brand-name drug have more of an impact on patients than the generic drug?

Result	Brand-name	Generic	Placebo
Improvement	24	21	10
No change	12	13	45

(a) Calculate the marginal frequencies.

(b) Find the expected frequency for each cell in the contingency table. Assume that the variables are independent.

Exercise 2. Is achieving a basic skill level in a subject related to the location of the school? The results of a random sample of students by the location of school and the number of those students achieving basic skill levels in three subjects is shown in the contingency table. At $\alpha = 0.01$, test the hypothesis that the variables are independent.

Location of school	Reading	Math	Science
Urban	43	42	38
Suburban	63	66	65

Class Exercise 2. The contingency table shows how a random sample of adults rated their local public schools and how they rated public schools nationally. At $\alpha = 0.05$, can you conclude that the adults' ratings are related to the type of school?

Type of school	Excellent	Good	Fair	Poor
Local	120	405	263	151
National	41	238	481	179

Definition: A **test of homogeneity**, we test the claim that *different populations* have the same proportions of some characteristics.

Exercise 3. Two different professors teach an introductory Statistics course. The table shows the distribution of final grades they reported. We wonder whether one of these professors is an “easier” grader.

	Prof. Alpha	Prof. Beta
A	3	9
B	11	12
C	14	8
D	9	2
F	3	1

Conduct a test of homogeneity.

When the sample data are from *one* population, we use a test of independence. When samples are drawn from different populations, we use a test of homogeneity.

Exercise 4. The table below shows the rank attained by male and female officers in the New York City Police Department. Do these data indicate that men and women are equitably represented at all levels of the department?

	Male	Female
Officer	21,900	4,281
Detective	4,058	806
Sergeant	3,898	415
Lieutenant	1,333	89
Captain	359	12
Higher ranks	218	10

(a) To see if there is evidence of differences in ranks attained by males and females, will you test homogeneity or independence?

(b) State the hypotheses.

(c) Test the conditions.

(d) How many degrees of freedom are there?

(e) Find χ^2 and the P-value.

(f) State your conclusion.

Class Exercise 3. A poll conducted by the University of Montana classified respondents by whether they were male or female and political party, as shown in the table. We wonder if there is evidence of an association between being male or female and party affiliation.

	Democrat	Republican	Independent
Male	36	45	24
Female	48	33	16

(a) To see if there is evidence of differences in ranks attained by males and females, will you test homogeneity or independence?

(b) State the hypotheses.

(c) Test the conditions.

(d) How many degrees of freedom are there?

(e) Find χ^2 and the P-value.

(f) State your conclusion.

Class Exercise 4. The following table shows the Myers-Briggs personality preferences for a random sample of 519 people in the listed professions (*Atlas of Type Tables* by Macdaid, MacCaulley, and Kainz). T refers to thinking and F refers to feeling.

	T	F
Clergy	57	91
M.D.	77	82
Lawyer	118	94

(#10) Use the chi-square test to determine if the listed occupations and personality preferences are independent at the 0.01 level of significance. **Answer: p -value = 0.0059**

Class Exercise 5. The following table shows ceremonial ranking and type of pottery sherd for a random sample of 434 sherds at a location in the Sand Canyon Archaeological Project, Colorado (*The Architecture of Social Integration in Prehistoric Pueblos*, edited by Lipe and Hengmon).

	Cooking Jar Shreds	Decorated Jar Shreds
A	86	49
B	92	53
C	79	75

(#12)

Use a chi-square test to determine if ceremonial ranking and pottery type are independent at the 0.05 level of significance. **Answer: p -value = 0.0451**

Homework

C Problems

9-19 ODD

B Problems

None

A Problems

None